Analysis of Mutational Drug Resistance of Tuberculosis by Comparing Multiple Tuberculosis Strains

Shihai Feng, Karina Yusim, Chang-Shung Tung, T-6; Tanmoy Bhattachary, T-2; Bette Korber, T-6

Finding the mechanisms of TB drug resistance is extremely important. In this study, we present a phylogenetic and computational analysis of five new, nearly complete TB genomes from KwaZulu-Natal, South Africa, together with 13 previously reported TB complete (or nearly complete) genomes from Genbank and Broad Institute. We found that strong phylogenetic relationships between TB samples from South Africa are critical for consideration and that the SNPs are not consistent across different XDR isolates.

Drug resistant tuberculosis (DR TB) is an increasing threat to the public health. In general, there are two types of DR TB, multi-drug resistant tuberculosis (MDR TB) and extensively drug resistant tuberculosis (XDR TB). MDR TB isolates display resistance to isoniazid and rifampicin and are associated with ~30% mortality. XDR TB isolates display the MDR profile with additional resistance to quinolones and one of the injectables, and are associated with 85% mortality. In 2008, there were an estimated 510,000 incident cases of DR TB with 270,000 deaths worldwide. DR TBs were identified in 114 different countries, of which 27 countries had a high burden of infection (World Health Organization).

There are several mechanisms through which TB can acquire drug resistance. These include mutations of DR TB through acquiring or losing genes and single nucleotide polymorphism (SNP) mutations. The identification of the mechanisms for drug resistance of TB is important to solving the global health problems that are related to diseases such as tuberculosis. Its implications for both national and global security are the reasons why health and bioscience research is an integral part of the LANL mission.

Here we present a phylogenetic and computational analysis of five new, nearly complete TB genomes from KwaZulu-Natal (KZN), South Africa, together with 13 previously reported TB complete (or nearly complete) genomes (one XDR TB, five MDR TBs, and six drug-sensitive TBs) from Genbank and Broad Institute. We focused on SNP mutation mechanisms. The basic steps we followed include: (1) assembling the contigs of the five new, nearly complete TB genomes from KwaZulu-Natal into full genomes, (2) grouping the homologous gene sequences from all strains and aligning them, (3) extracting SNPs, in which at least one of 18 sequences is different from others, from all 18 strains of which each has about four million nucleotide bases, and (4) making molecular parsimony trees by comparing artificial sequences that contain only the SNPs in each stain.

As shown in Fig. 1, there is a strong correlation of the year of TB diagnosis with the distance from the most recent common ancestor of KZN lineage and a possible co-infection of XDR patients from Tugela Ferry and Durban with DS and MDR strains clustering with F11 and Bejing—all XDR TB strains in KZN form a sub-lineage. Further studies (Fig. 2) based on comparing non-synonymous (a mutation in a nucleotide in the DNA sequence that results in an amino acid change in the protein) and synonymous mutations (a mutation in a nucleotide in the DNA sequence that does not result in an amino acid change in the protein) reveal that the XDR strains in KZN lineage are a shared lineage that is independent of drug resistance. Comparing the strains DS_KZN_4207 to MDR_KZN_1435, MDR_KZN_V3475, and four XDR KZN strains, we found a set of 55 non-synonymous SNPs. None of 55 non-synonymous SNPs were found in the set of SNPs observed in the comparison of DS_KZN_17030 to MDR_KZN_813 strains. Furthermore, we found that the rpoB and katG genes have distinct mutations in both sets. It is not only the correlated SNPs, but also the mutations in key genes/proteins that have the most importance for drug resistance.

For more information contact Shihai Feng at sfeng@lanl.gov.

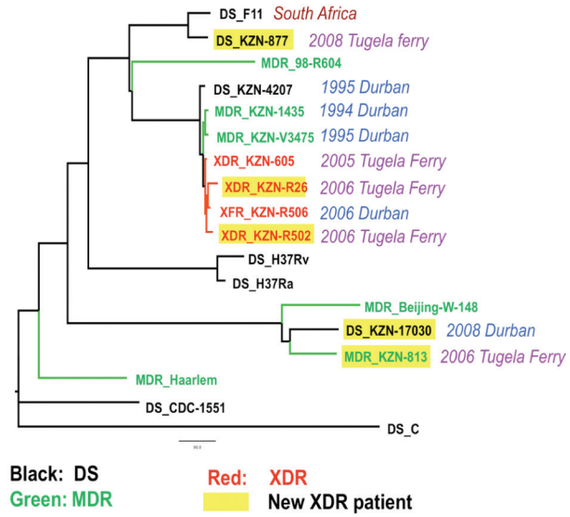
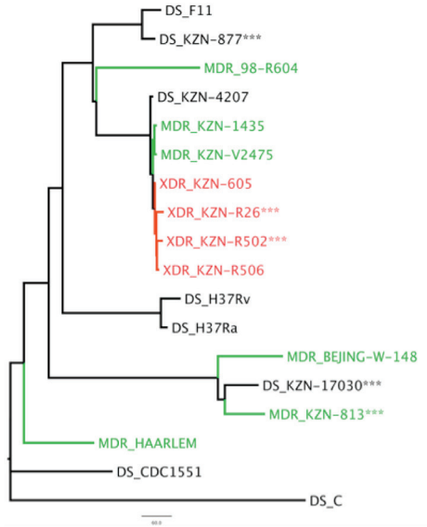
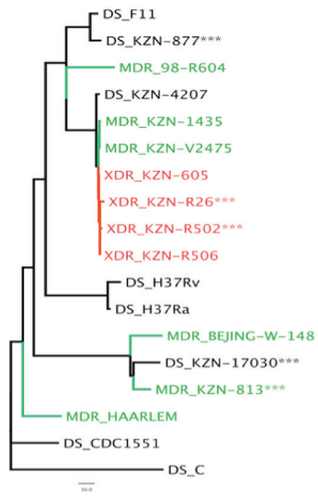


Fig. 1. The parsimony tree based on 4,586 variable positions from protein-coding regions.



Non-synonymous tree



Synonymous tree

Fig. 2. The parsimony trees based on synonymous SNPs and non-synonymous SNPs from protein-coding regions are virtually identical, indicating that branching order is mostly defined by founder effect rather than by drug resistance.

[1] Gueron, M., et al., *Nature* **328,** 89 (1987).
[2] Alexandrov, B.S., et al., *Nucleic Acids Res* **38,** 1790 (2010).
[3] Alexandro, B.S., et al., *J Biol Phys* **35,** 31 (2009).
[4] Alexandrov, B.S., et al., *PLoS Comput Biol* **5,** e1000313 (2009).
[5] Alexandrov, B.S., et al., *Phys Lett A* **374,** 1214 (2010).
[6] Bock, J., et al., *PLoS ONE* **5,** e15806 (2010).

Funding Acknowledgments
LANL Laboratory Directed Research and Devlopment